USGS Community for Data Integration Data Management Working Group: Science Data Catalog, Open Data Discussion

November 10, 2014

Goals

1. Background on Administration Open Data requirements, DOI implementation, etc.
2. Overview of USGS Science Data Catalog (SDC), pathways to data.doi.gov, data.gov, and the Geospatial Platform
   a. Providing records to SDC, data.doi.gov, data.gov and the geoplatform
3. Inventory and status of NGDA and A16 datasets (NGDA and A16 leads)--tracking
4. Next steps (all)

USGS

Background & Drivers…

1. On May 9, 2013, the White House released the Executive Order, "Making Open and Machine Readable the New Default for Government." This Order, built upon an earlier interagency memorandum released by the OSTP, was accompanied by an OMB Policy requiring:

   - ❑ Inventory of Data
   - ❑ Public Data Listing
     - ❑ data.doi.gov
     - ❑ data.gov

2. USGS Science Data Management activities

   - ❑ Community for Data Integration
   - ❑ Office of Science Quality and Integrity – policy revisions
   - ❑ Moving toward a "science data life cycle approach" in USGS
   - ❑ Implementation of USGS Science Strategy Plans

USGS

Background & Drivers cont…

Helping address Big Earth Data Initiative (BEDI) actions:

  * Making large, complex datasets accessible, machine actionable
  * Treating these datasets to improve interoperability, accessibility
  * Coordinating across DOI, and with NASA, NOAA

≋ USGS

Stakeholders & Project Goals…

Outreach & Communications

Publishing

Infrastructure & Assurance

Quality & Integrity

Project Goals:

- **Least impact to all**
- **Self-service model**
- **Broad involvement**
- **Increase Program visibility**
- **Proper attribution**
- **Incremental updates**
- **Improve Data Mgmt.**
- **Effective links to related efforts**
- **Reuse of Standards**
- **Leverage existing systems, efforts**
- **Move towards actionable data**

≋USGS

What is the USGS Science Data Catalog (SDC)?

1. Single point for USGS to provide its metadata to data.doi.gov, data.gov, OMB, geoplatform etc.
2. Organizes USGS data – **find, get, use (Actionable)**
3. Meets Open Data requirements for USGS
4. Provides metadata retrieval, visualization, download, & links back to original providers

≋USGS

What is the USGS SDC (cont)?

1. Allows data stewards control of what data to publish and review metadata for release – through the "SDC Dashboard"

2. Requires metadata content prior to data release--FGDC CSDGM, no broken links (See http://sunlightfoundation.com/blog/2014/10/30/federal-open-data-audit-defense-downright-dismal-interior-immense-yet-imperfect/ for an example of the perils of broken links…)
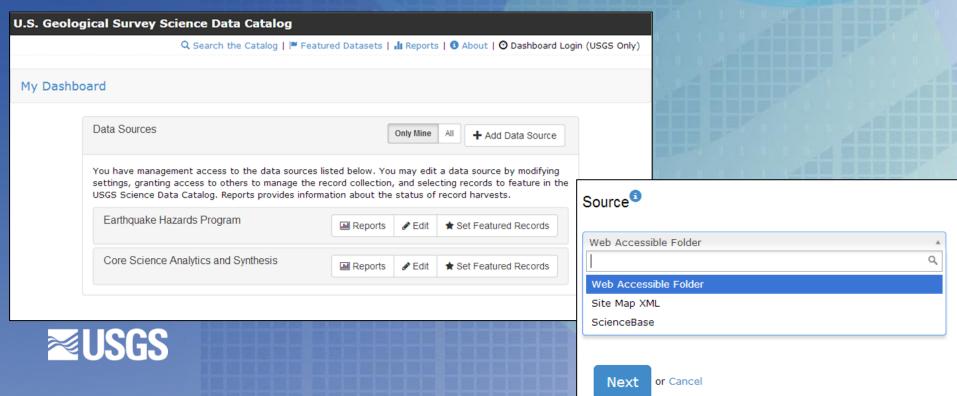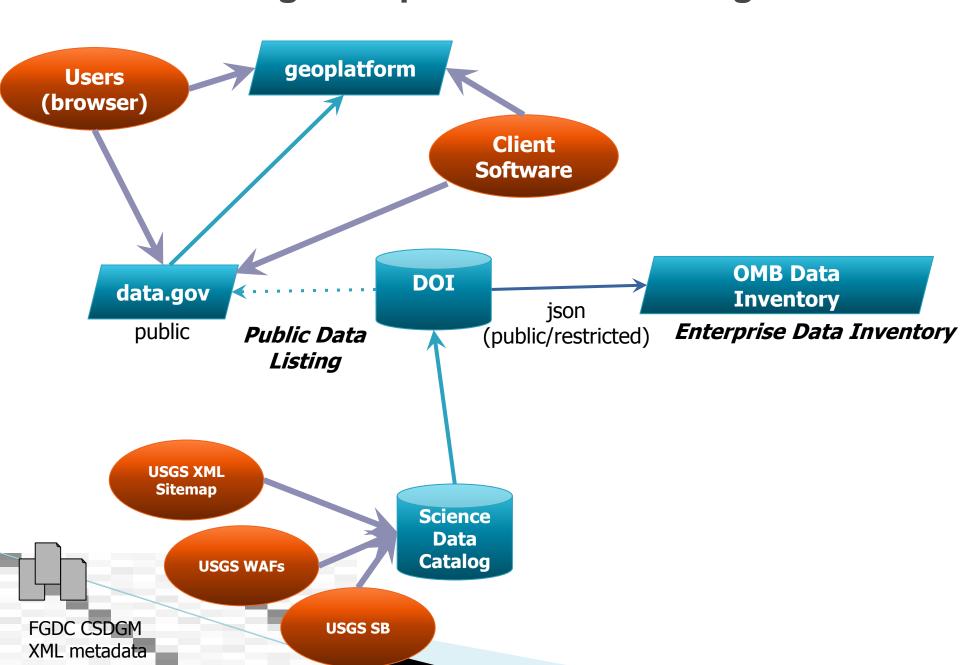
3. Focus on scientific data

USGS

Science Data Catalog (SDC) Dashboard

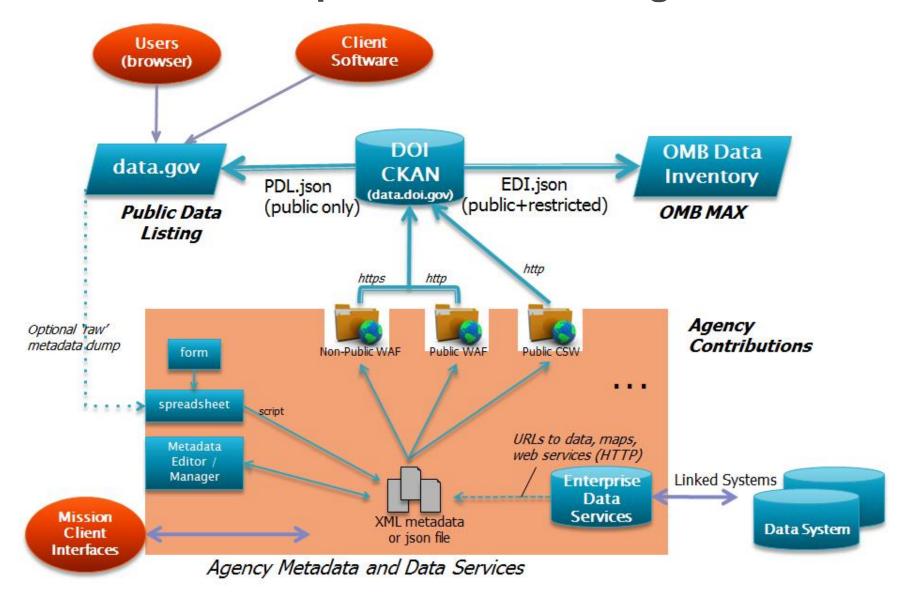**Allows managers and data stewards to:**

* Register/Define/Maintain metadata sources
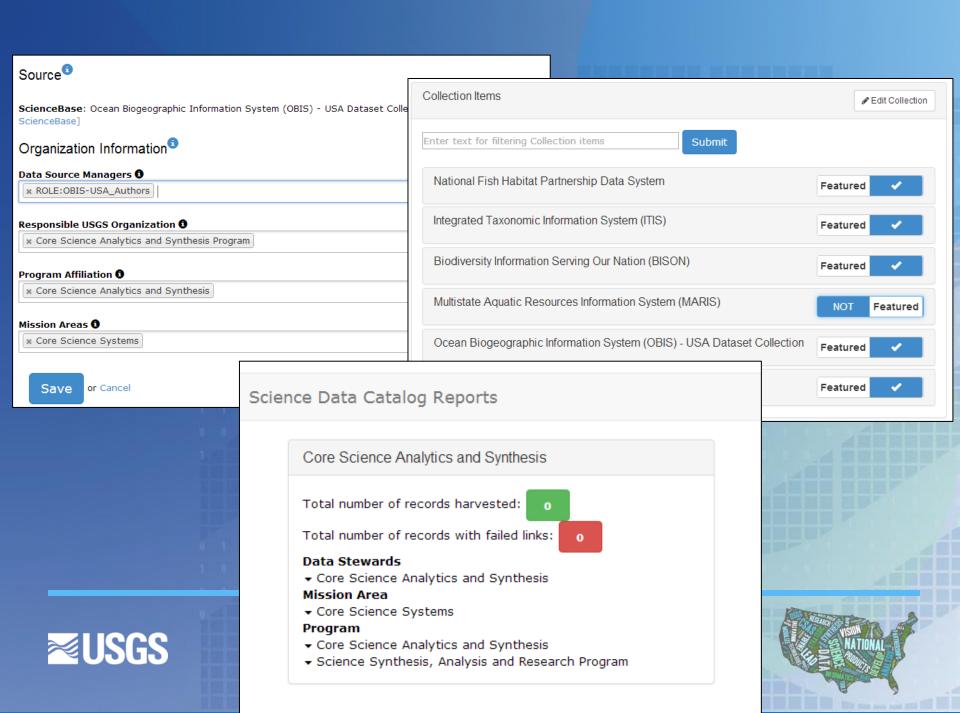* Obtain statistics on metadata indexing, completeness, and errors

Data Publishing for Open Data and data.gov
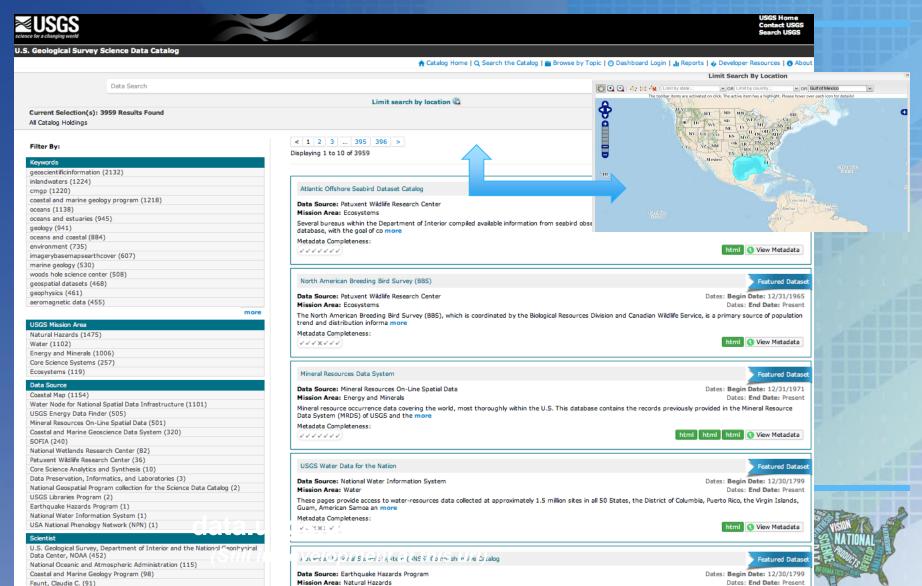
Data Release for Open Data and data.gov

Source ⓘ

ScienceBase: Ocean Biogeographic Information System (OBIS) - USA Dataset Colle
ScienceBase]

Organization Information ⓘ

Data Source Managers ⓘ
× ROLE:OBIS-USA_Authors

Responsible USGS Organization ⓘ
× Core Science Analytics and Synthesis Program

Program Affiliation ⓘ
× Core Science Analytics and Synthesis

Mission Areas ⓘ
× Core Science Systems

Save   or Cancel

Collection Items   ✎ Edit Collection

Enter text for filtering Collection items   Submit

National Fish Habitat Partnership Data System   Featured ✓

Integrated Taxonomic Information System (ITIS)   Featured ✓

Biodiversity Information Serving Our Nation (BISON)   Featured ✓

Multistate Aquatic Resources Information System (MARIS)   NOT Featured

Ocean Biogeographic Information System (OBIS) - USA Dataset Collection   Featured ✓

Featured ✓

Science Data Catalog Reports

Core Science Analytics and Synthesis

Total number of records harvested: 0

Total number of records with failed links: 0

Data Stewards
▾ Core Science Analytics and Synthesis
Mission Area
▾ Core Science Systems
Program
▾ Core Science Analytics and Synthesis
▾ Science Synthesis, Analysis and Research Program

≋ USGS

USGS Science Data Catalog-- data.usgs.gov

DOI -- Data.doi.gov

Data.gov

Geospatial Platform

USGS Science Data Catalog Integration with New USGS Web…

- **Partnering with OCAP and AEI**
- **Drive "data" piece of new USGS Web**
- **Minimizes duplicate effort**

SDC Metadata Data Stewards & Sources…

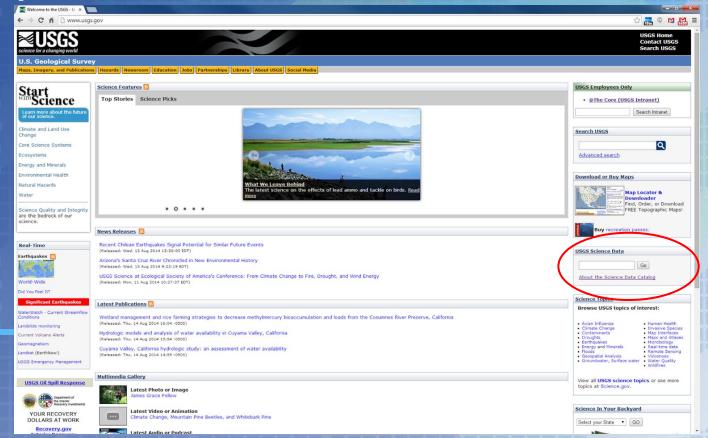| Source Name	Organizational Unit	Mission Area	Point of Contact	Catalog Status	System-level Record or Individual Records?	"Featured Dataset"?	Current Total # of Records Passing Validation in the Catalog
Mineral Resources Data (MRData)	Minerals Program	Energy & Minerals and Environmental Heath	Peter Schweitzer	In Catalog	BOTH	Yes, for system-level record	501
Patuxent Wildlife Research Center	Patuxent Wildlife Research Center	Ecosystems	Mark Wimer	In Catalog	Individual Records	Yes, for 2 records: Breeding Bird Survey & Atlantic Seabird Database	35
US Topographic Maps	National Geospatial Program	Core Science Systems	Rick Brown/Brian Fox	In Catalog	System-level Record	Yes	1
Historical Topographic Maps	National Geospatial Program	Core Science Systems	Rick Brown/Brian Fox	In Catalog	System-level Record	Yes	1
National Geologic Map Database	National Cooperative Geologic Mapping Program	Core Science Systems	Dave Soller	Not Yet Started	System-level Record	Yes	not yet started
Landsat	EROS	Climate & Land Use Change	John Faundeen/Wayne Miller	In Process	System-level Record	Yes	in progress
Advanced National Seismic Systems (ANSS)	Earthquakes Program	Natural Hazards	Harley Benz, ANSS Technical Manager, (benz@usgs.gov); Cecily Wolfe, ANSS Coordinator, (cwolfe@usgs.gov);	In Catalog	System-level Record	Yes	1
USGS Water Data for the Nation	National Water Information System	Water	Dorrie Gellenbeck	In Catalog	System-level Record	Yes	1
National Hydrography Dataset	National Geospatial Program	Core Science Systems	Jeff Simley (sp?)	Not Yet Started	System-level Record	Yes	not yet started
Did You See It?	Landslides Program	Natural Hazards	Rex Baum?	Not Yet Started	System-level Record	Yes	not yet started
NAWQA Data Warehouse	CIDA	Water	Nate Booth	Not Yet Started	System-level Record	Yes	not yet started
Integrated Taxonomic Information System (ITIS)	Core Science Analytics & Synthesis	Core Science Systems	Gerald "Stinger" Guala	In Catalog	System-level Record	Yes	1
Ocean Biogeographic Information System (OBIS-USA)	Core Science Analytics & Synthesis	Core Science Systems	Mark Fornwall	In Catalog	System-level Record	Yes	1
Biodiversity Information Serving Our Nation (BISON)	Core Science Analytics & Synthesis	Core Science Systems	Gerald "Stinger" Guala	In Catalog	System-level Record	Yes	1
Gap Analysis Program (GAP)	Core Science Analytics & Synthesis	Core Science Systems	Kevin Gergely	In Catalog	System-level Record	Yes	1
GAP Species Ranges	Core Science Analytics & Synthesis	Core Science Systems	Kevin Gergely	In Catalog	System-level Record	Yes	1
GAP Species Distribution Models	Core Science Analytics & Synthesis	Core Science Systems	Kevin Gergely	In Catalog	System-level Record	Yes	1
GAP Land Cover Data	Core Science Analytics & Synthesis	Core Science Systems	Kevin Gergely	In Catalog	System-level Record	Yes	1
National Fish Habitat Partnership Data System (NFHP)	Core Science Analytics & Synthesis	Core Science Systems	Andrea Ostroff	In Catalog	System-level Record	Yes	1
Protected Areas Database of the United States (PADUS)	Core Science Analytics & Synthesis	Core Science Systems	Kevin Gergely	In Catalog	System-level Record	Yes	1
Multistate Aquatic Resources Information System (MARIS)	Core Science Analytics & Synthesis	Core Science Systems	Andrea Ostroff	In Catalog	System-level Record	Yes	1
USGS Photographic Library	USGS Libraries Program	Core Science Systems	Keith Van Cleave	In Catalog	System-level Record	Yes	1
USGS Publications Warehouse	USGS Libraries Program	Core Science Systems	Helen Tong	In Catalog	System-level Record	Yes	1
USA National Phenology Network (NPN)	Status & Trends Program	Ecosystems	Jake Weltzin	In Catalog	System-level Record	Yes	1
Geomagnetic Observatories	Geomagnetism Program	Natural Hazards	Duff Stewart	Not Yet Started	TBD	TBD	not yet started
Coastal & Marine Geoscience Data System (CMDGS)	Coastal & Marine Geology Program	Natural Hazards	Greg Miller	In Catalog	Individual Records	No	323
CoastalMap	Coastal & Marine Geology Program	Natural Hazards	John O'Malley	In Catalog	Individual Records	No	1167
InfoBank	Coastal & Marine Geology Program	Natural Hazards	Rex Saunders	Not Yet Started	Individual Records	No	not yet started
National Wetlands Research Center	National Wetlands Research Center	Ecosystems	Cassie Thibodeaux	In Catalog	Individual Records	No	82
Energy Data Finder	Central Energy Resources Team	Energy & Minerals and Environmental Heath	Greg Gunther	In Catalog	Individual Records	No	505
Water NSDI Node	Water HQ	Water	Mike Ierardi	In Catalog	Individual Records	No	1101
South Florida Information Access (SOFIA)	Priority Ecosystem Science	Ecosystems	Heather Henkel	In Catalog	Individual Records	No	240
SageMap	Snake River Field Station, FRESC	Ecosystems	Tom Zarriello	Not Yet Started	Individual Records	No	not yet started
USGS Core Research Center Core Collection	Data Preservation, Informatics, and Laboratories	Core Science Systems	Betty Adrian	In Catalog	System-level Record	No	1
USGS Core Research Center Cuttings Collection	Data Preservation, Informatics, and Laboratories	Core Science Systems	Betty Adrian	In Catalog	System-level Record	No	1
National Ice Core Laboratory	Data Preservation, Informatics, and Laboratories	Core Science Systems	Betty Adrian	In Catalog	System-level Record	No	1
National Atlas	National Geospatial Program	Core Science Systems	Brian Fox ?	Not Yet Started	?	?	not yet started

As of February 20, 2014

Contacts

Ben Wheeler: **bwheeler@usgs.gov**

Links:
SDC: data.usgs.gov
SDC dashboard: www.sciencebase.gov/catalog/dashboard
DOI CKAN: data.doi.gov
data.gov: data.gov
Geospatial Platform: geoplatform.gov

- **USGS Science Data Catalog Team: Mike Frame, Viv Hutchison, Natalie Latysh, Lisa Zolly, Michelle Chang**

≋ **USGS**